Advances and Applications in Mathematical Sciences
Volume 20, Issue 10, August 2021, Pages 2095-2112

TWITTER BASED TRAFFIC ANALYSIS AND PREDICTION MODEL FOR PLANNED EVENTS

HARSHA S. RATNANI and SIDDHARTH KUMAR

Department of Information Technology
JIMS, VK, New Delhi, India
E-mail: harshapratnani@gmail.com
Indiasiddharth.kumar2506@gmail.com

Abstract

This paper has been written with the fact the study of extracting relative information from social media for long term and short term events and finding its link with traffic related management and control or finding better ways to manage huge traffic jams. As a newly emerged communication revolution, Day to day lives of people is now often connected to their social media accounts and is observed to be a platform for users to communicate, share, and follow the interesting things happening in their daily lives in an instantaneous channel. The number of posts posted online on any platform such as Twitter, WhatsApp or Facebook, related to any social event can somewhat represent its corresponding attention levels. We focused on tweets posted on Twitter for events, in which we used tweets of social events involving some kinds of trip requirements to and fro from the venue, as it usually leads to an obvious traffic increase in the surrounding area. To prove the correlation between twitter semantics and traffic conditions, our study focuses on using the tweets related to sporting games to predict the passenger flow, which is strategically important in metro transit system management.

I. Introduction

Often nowadays whenever we go for any major social event, the first hurdle we all have to face is the big traffic jam specially surrounding the venue, and especially during the start and end of the event on the way to that leads up to the venue. The general traffic operations may deteriorate around major social events including ceremony opening, celebrity death, festival parades, international conference, etc. Such major events across the world are vastly covered on social media platforms such as twitter.

2010 Mathematics Subject Classification: 62H11, 62H30, 62H35, 62M10, 62M30, 26E70, 83C75.
Keywords: Social Media, Traffic, Events, Twitter, Hashtag, Tweets, Prediction, Regression, Analysis.
Received October 7, 2020; Accepted January 15, 2021

It is observed that with expanding number of engine vehicles across the world, street traffic prediction is becoming necessary day by day and is has become critical component in modern smart transportation systems. Accurate and exact prediction of both the short-term and longer-term street traffic conditions can greatly help metro cities traffic management agencies in planning proactive strategies to handle the congestions on the street during peak hours. Not only that It can also help travelers to plan their trips accordingly either by leaving early or by avoiding those routes completely which are expected to be congested soon with huge traffic jam.

The social media platform Twitter has provided us with recently progressive technique for information diffusion, and this colossal volume of messages, data and information by Twitter has excited the interests in many researchers from various fields such as opinion polls, geographic data study, urban smart systems planning, audience movie reviews, etc. The majority of these research works have demonstrated promising outcomes which both upgrade the conventional systems and widen the new research spectrums. Motivated by the magnitude of the information present in online social media, in this paper, we try to analyse if we can use tweet-based semantics to provide clues about the traffic condition during the occurrence of a major event? To attain this, we take into consideration two types of analysis-Long Term and Short Term. Long term analysis focuses on taking twitter and traffic data for a long period (5 to 10 years), during the yearly occurrence of any major event, and find the correlation among them. Whereas the short term analysis takes into consideration the same data during the short span of time i.e., during the occurrence of the event. The duration may vary from single day to a week with respect to a given year.

In this study we took up the case of one of the most popular events in the world - The Wimbledon Tennis Championship, along with the case of football matches played by Manchester United at Old Trafford over a span of 10 years, for long term analysis. On the other hand, we have considered the case of a single day match of the Indian Premier League 2017 for the short term analysis during the day of the match between two teams. We found that tweet semantics can be a positive indicator for traffic judgment.

The rest of the paper has been organised as follows. In Section two, we intended to briefly review the prevailing research work on traffic prediction

and data analysis based on social media aided gathered data. Then we have shown the study of our proposed methodology for both types of analysis long term and short term planning in Section three. The experimental results of the data which have gathered from specific case studies are presented in Section four. Finally, we tend to conclude the paper in Section five and have listed out all the references for this paper in Section six.

II. Referenced Work

Social Media Based Analysis

It's been observed by many researchers that the rich information available on social media platforms such as twitter, facebook, whatsapp etc. can be utilized for various application and data analytics purposes. As we all have seen recently, there is a lot of interest in using social media to detect emerging news or events: in Petrovic et al., [1], the researchers have addressed the problem of detecting new events or first story detection from a stream of Twitter posts using an algorithmic locality-sensitive hashing approach. They used method of adapting to the first story detection task by introducing a back off towards exact search. As claimed in the paper this adaptation greatly improved performance of the system and virtually eliminated variance in the results. They used this FSD system on a large-scale task of detecting new events from millions of Twitter posts;

Then in Sakaki et al., [3], the authors had investigated the interaction during natural calamities events such as earthquakes on Twitter, and proposed a probabilistic spatiotemporal model for the target event that can find the center and the trajectory of the event location, it was an application for earthquake reporting etc. They applied the semantic analyses to tweets to classify them into a positive and a negative class. They considered each Twitter user as a sensor, which can help in detecting an event based on sensory observations. They used in their research Location estimation methods such as Kalman filtering and particle filtering to estimate the locations of events. In Sankaranarayanan et al., [2], the authors proposed a news tweets processing system called to capture tweets that correspond to late breaking news, they named it as Twitter Stand. They used naïve Bayesian classier to improve the quality of the noisy feeds and employed a

dynamic corpus to sensitize the classier to current news. They observed that the sheer enormity of the data means that algorithms will have to be online in nature, which can be challenging. The online clustering algorithm that they presented in their paper was useful along with being fast and robust for mitigating noise. In addition, they also described methods for geotagging news, as well as a user-interface for displaying news.

The other line of research is tweet classification focused on information filtering. It was found that in Go et al., [4], the authors test various algorithms for classifying the sentiment of tweets, such as SVM, Naive Bayes, etc based on author information and some other features within the tweets, it is known Bag-of-Words approach within the tweets. With such a system, users could subscribe to or view only certain types of tweets based on their interest. Though the approach didn't work with lot of noise into the data, hence any noise removal techniques needs to be applied first on are necessary in such cases; in Sriram et al.,[5], the researching team used a small set of domain related features along with the bag-of-words features to describe and then classify the tweets into a predefined set of classes; etc. show that changes in the public mood state can indeed be tracked from the content of large-scale Twitter feeds by means of rather simple text processing techniques and that such changes respond to a variety of socio-cultural drivers in a highly differentiated manner.

During our literature study and moving more in time, we discovered some group of analysts and researchers were extracting information from tweets which might be useful in another domain. Like in Bollen et al., [6], the authors tried to perform the research based correlation between public mood and other economic indicators. So in their research they started deriving collective mood states from the large scale Twitter Feeds and then performed the correlation analysis with the Dow Jones Industrial Average (DJIA) over time. Finally, they had concluded that the accuracy of DJIA by the inclusion of specific public mood dimensions, such as Calm the predictions could be significantly improved. Also, we found that in Eisenstein et al., [7] the authors proposed a multi-level generative model which was based on the geo-tagged social media, which brought the reasons jointly about latent topics and geographical regions into the light.

Road Traffic Prediction

In metros and otherwise big urban cities Traffic Prediction is an important as well as a critical component in any smart transportation systems. It's possible that if we can do the accurate prediction of traffic conditions, then this would help traffic management agencies to plan and handle the city traffic congestion problem with a proactive traffic operation strategy for avoiding it at first place and also help them to efficiently handle these congestions on roads knowing well in advance along with the common road travelers can also plan their trips accordingly ahead of time in case such warnings are issued publicly.

It came to our notice that studies based on long-term events based traffic prediction are rather very limited, essentially in light of the fact that extra factors other than the past and current traffic conditions start to play a very important role once the forecasting time period is beyond 60 minutes. We found out that only group of few researchers and private sector companies have attempted to analyse and utilise the correlation between the route traffic data and the other external factors such as weather and event schedules Maze et al., [8]; Mahmassani et al., [9].

Our proposed work focuses on analysing the correlation between Twitter and Traffic in context to a particular well known major event, and further predicting the traffic in future based on this analysis. This is inspired and motivated by observing the large occurrence of chats, posts and tweets shared on social media platform which are directly related to traffic conditions, as mentioned in Ni et al., [11], which describes forecasting subway passenger flow during event occurrences, and Ozdikis et al., [12], which further discusses about event detection based on the use of hashtags in twitter.

III. Proposed Methodology

In this proposed research study related to various public events data, we mainly used two kinds of data items for the study: the tweets and the traffic data. The first of them both we collected through Twitter Streaming with geo-location filter. To keep the study focused on fixed public events all tweets are paired with time and location information. The Traffic data is then extracted using the Google Maps to analyse the correlation between tweets and traffic conditions.

Long Term Analysis

During our study we first focused on long term data analysis just so that the certain event occurrences have a possibility of getting being affected by any unforeseen circumstances like any occurrence of an accident on the route or may be due to change of weather like heavy rainfall etc. So in prediction of the effects of tweets and its correlation with traffic in any place, we thought we must consider a long term data first (over a span of at least 10 years) for the study.

In this study we took up the case studies of two of the most popular events in the world - The Wimbledon Tennis Championship, and the football matches played by Manchester United at Old Trafford over a span of 10 years. The reason for these events being chosen is because of their invariability. The matches in Wimbledon take place at the same time each year, and continue for the exact same period of time. The tweets of this event have limited hashtags and are contained within or around the word "Wimbledon", and for Manchester United, the hashtags are contained mostly around "GGMU". Hence, search of tweets with these words arguably will contain all tweets related to the event.

While our study has considered the case of the football matches played by Manchester United to study the direct relation between tweet semantics and traffic data, the case of Wimbledon has been considered further to predict the amount of tweets in the future consisting of #Wimbledon, and then predict the amount of traffic on the roads being taken into consideration.

To accommodate the actual correlation analysis as per our suggestive model, we used two data record sets: the one was containing the traffic measurements, and the other one had the gathered tweets posted, during the period of the event. We generated the traffic data set by collecting measurements for two specific roads around the Wimbledon Stadium - B235 and Riverside Road. The cumulative traffic data for these two roads was obtained from 2006-2016, during the period when the Wimbledon Championship takes place. Using the geo-location filters based on latd/long bounding box on the gathered data, we obtained all the relative tweets.

After the data obtained from maximum number of attempts using twitter was gathered, we were able to figure out the number of tweets during that

duration that contained #Wimbledon in it. To avoid any kind of spam, we also took care of applying a filter on the tweets that contain the regular expressions of "http:" or "www.". For each tweet, we collected the information of respective user account, the tweet time stamp, the tweet content and the geo-location of the twitter user.

In order to figure out the trend of twitter popularity on events such as Wimbledon, the long term pattern needs to be plotted. The different values of the tweet concentration for various years are then analysed to predict a pattern. If a pattern is found, the mathematical analysis of the pattern can help in predicting future values. In the case of Wimbledon, the tweet concentration followed a linear pattern. This gave rise to a simple linear equation of $Ax + Bx + C = 0$ type. If not linear, the tweet concentration would've followed a non-linear pattern of the type $y = 3x^3 + x^2 - 7$. Using this equation the concentration of tweets in the future years was predicted, which could be used as an input in the traffic prediction of the future. The data set collected showed a linear pattern in tweet concentration making it easy to predict the value of the future.

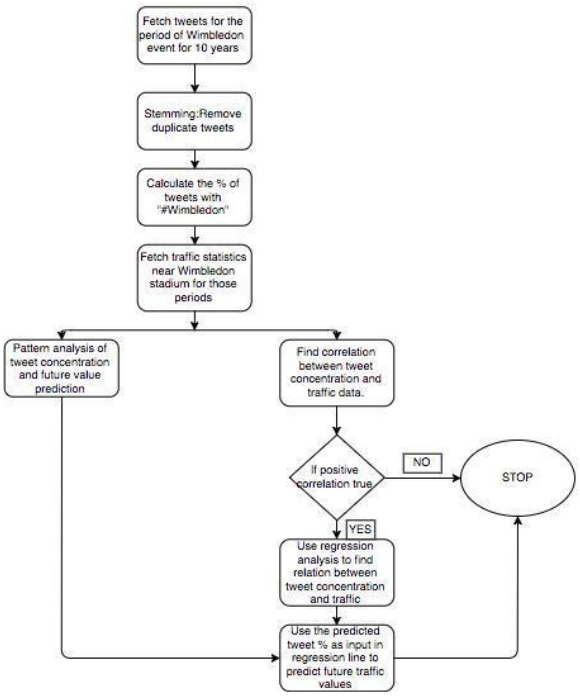
The correlation between tweets and actual traffic data needs to be mapped. Taking the data of past years, the relation between the tweet concentration of any particular event (Wimbledon) can be figured out by mathematical linear regression tool. This allows to take the data of the two distinct variables, find a correlation coefficient between them, and thus predict a mathematical equation that can be used to plot the values of each year, and extend this line into the next few years to predict one variable using the other. In this case the expected tweet concentration value was plugged in to predict what the traffic in the nearest roads to the event may look like sometime in the future.

Short Term Analysis

In contrast to the above analysis, the short term analysis considers the twitter and traffic data over the period of the occurrence of the event, on an hourly basis. In our study, we have considered the a cricket match of the Indian Premier League, that lasted for a single day, and had multiple variations of the amount of tweets and traffic during that particular duration.

For studying the short term analysis of the effects of tweets posted on any kind of traffic situation at any particular place, we developed a hashtag based event search algorithm. In order to collect the tweets which are being posted by Indian users, we first defined geographic bounding boxes that actually covers almost all of India, and added the same as our filtering criteria for the streaming service. The lat/long bounding box of "23.546471, 78.981548, 3000 km" is used to get tweets from this location, i.e., Madhya Pradesh, with a radius of 3000 km around it.

By extracting tweets based on their location, we further determine the traffic at that particular location using the traffic and transit layers of Google Maps. Google Maps allows you to add real-time traffic information to your maps using the Traffic Layer object. We store this traffic information for there quisite location with a timestamp in a SQL Database, and plot it on different graphs to analyse the correlation between the event in concern and the traffic at various hours of the day, which can be further used to accordingly prepare traffic management plans in case of other events in the same area in the future.



IV. Experimental Results

Long Term Analysis

United Kingdom is equipped with sensors across all major roads collecting data on traffic each day. This traffic data is stored in the UK Department of Traffic database and is accessible to anyone across the globe. The sensors store data on various parameters such as, average traffic per day, traffic for the year, traffic of different vehicle types etc. Such data is extremely valuable in data analysis and is a very helpful tool. By crawling through official site (https://www.dft.gov.uk/traffic-counts) and filtering data with respect to geo-location and time, exact traffic data was obtained and used directly in further analysis.

The graphs below (Figure 2 and Figure 3) show the number of motor vehicles across UK for 10 years. Similarly, information for every individual road can be extracted from the Department of Traffic database, and be plotted for better understanding of the correlation between tweets and traffic.

After extracting the traffic for the roads - B235 and Riverside Road, for the years 2006-2016, the tweets containing #Wimbledon were extracted for all of these years, and there spective graphs (Figure 2 and Figure 3) showed clear correlation between the number of tweets and corresponding change in traffic. The year of concern was tagged during extraction of twitter data, and all tweets were extracted until twitter blocking. This process was repeated multiple times and a compilation of all collected tweets from that year was made. Repetitive tweets were removed. The percentage of tweets that had them gave us the final tweet concentration which could then be used in further regression analysis.

Until 2010, the data was not sufficient to obtain a clear pattern or trend, i.e., the data was below the requisite confidence level. However, after 2010, the data showed a clear relation between the traffic in the two roads and the percentage of tweets consisting of #Wimbledon.

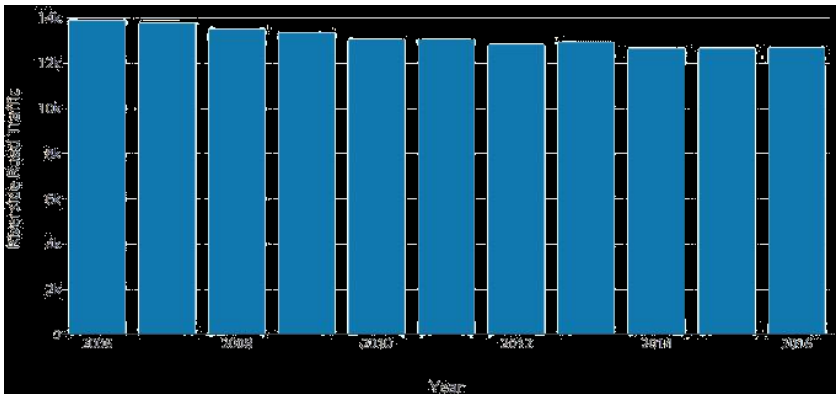


Figure 2. Amount of traffic on B235 from 2006-2016 during the Wimbledon Championship.

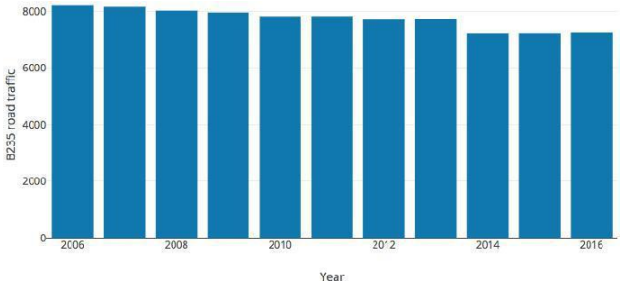


Figure 3. Amount of traffic on Riverside Road from 2006-2016 during the Wimbledon Championship.

Based on the above information, we plot the percentage of tweets consisting of #Wimbledon, during the period of the event, across 10 years, i.e., 2006-2016, with the amount of traffic, as shown. We observe that there is a positive correlation between the percentage of tweets and amount of traffic. Whenever the percentage of tweets consisting of #Wimbledon falls down, or rises in a particular year, there is a corresponding reduction or increase in the amount of traffic on the streets being considered, as can be seen in the graph shown below.

Due to lack of twitter usage during the years 2006, 2007, and 2008, the amount of data requisite for analysis was inadequate, and hence the percentage of tweets consisting of #Wimbledon has been taken 0 for this period.

Table 1. Tweets per year.

Tweet Perce	Year
0	2006
0	2007
0	2008
0.01	2009
0.03	2010
0.02	2011
0.07	2012
0.05	2013
0.04	2014
0.05	2015
0.06	2016

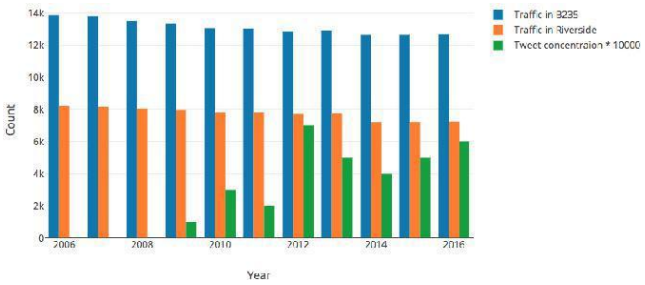


Figure 4. Tweet percentage per year (2006-2016) containing #Wimbledon (for the duration of the event) mapped to the traffic across the roads B235 and Riverside for the same duration.

To further analyse the correlation, we picked up the case of another event - a football match at Old Trafford Stadium in Manchester, London. Our analysis showed that with rise or fall in the number of tweets, the corresponding traffic for any particular year increased or decreased accordingly. #GGMU was used to extract tweets, which is the most used hashtag for the team - Manchester United, which plays its matches at Old Trafford.

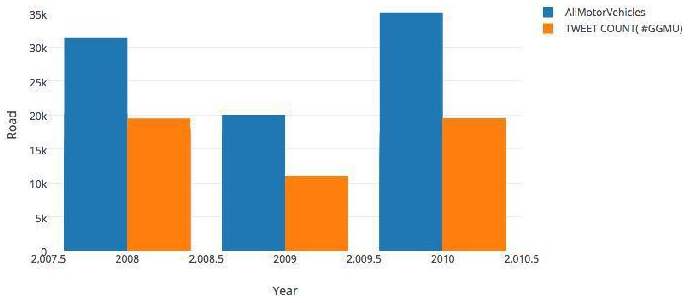


Figure 5. Number of tweets containing #GGMU and corresponding traffic in Manchester.

The main purpose of understanding the relation between tweets and corresponding traffic is to improve traffic management by prediction of traffic on particular roads for the future. Following steps are performed for the regression analysis of the case considered for #Wimbledon and roadsB235 and Riverside Road :

Step 1. Prediction of the percentage of tweets consisting of #Wimbledon in 2017, based on previously extracted data for10 years (2006-2016) for the duration of Wimbledon:

We calculate the percentage of tweets (y) consisting of #Wimbledon, where x is the index of the year of the tweet, when 10000 tweets are extracted for the period of 10 years, i.e., 2006-2016, during the duration of the event (Wimbledon) each year, in the equations given below:

$$2014: y = 0.04\%(x = 1) \tag{1}$$

$$2015: y = 0.05\%(x = 2) \tag{2}$$

$$2016: y = 0.06\%(x = 3) \tag{3}$$

Using the obtained values mentioned above, we computed the type of correlation (linear, exponential or haphazard), and used the following mathematical equation:

$$y - Mean(y) = r^*(\text{Variance}(y)/\text{Variance}(x)) \tag{4}$$

where

$$\Gamma = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum y)^2][n\sum y^2 - (\sum y)^2]}}.$$

Mean $(y) = 0.05$ (From (1), (2) and (3))

Variance $(y) = 0.01$ (From (1), (2) and (3))

Variance $x = 1$ (From (1), (2) and (3))

to give

$$y = 0.024x.$$

From the obtained equation, we inferred that tweets follow a linear regression model and the percentage ratio of tweets consisting of #Wimbledon should be 7% in 2017.

Step 2. Prediction of the amount of traffic on the roads B235 and Riverside Road in 2017, based on the analysis in step 1, and previously available traffic data provided by the UK Department of Traffic :

Further, to predict the amount of traffic in 2017, in accordance with the percentage of tweets, we used another regression model. The traffic on the two roads for the three different years, for the duration of the event Wimbledon, is as given below in table 2:

Table 2. Number of vehicles on Riverside Road and B235 for three years.

Year	Riverside Road	B235
2014	12659	7212
2015	12660	7214
2016	12687	7239

Mean of Traffic for Riverside Road =12668.67 (5)

Mean for Traffic for $B235 = 7221.67.$ (6)

Mean of Tweets concentration = 0.05

(From (1), (2), and (3)).

We then substitute the values calculated above into equation (4), to get the following linear equations for prediction:

B235 :

$$y = 1400x + 12599 \quad (7)$$

Riverside Road:

$$B235 : y1350 + 7154. \tag{8}$$

In accordance with the equations above, we insert the values calculated in equations (5) and (6) into (8) and (7) respectively. Based on this calculation, we can easily predict that the number of motor vehicles on Riverside Road for the duration of Wimbledon 2017 would've been 12697; whereas for B235 Road, the number of vehicles would've been 7248.

Short Term Analysis

In the preprocessing step, we are supposed to toke nise the tweeted sentences into words by using space and punctuations as separators or stop words to be specific. After stop words are eliminated as much as possible we then remove the non-alpha numeric characters within these tweets. The tweets are required to be extracted using a particular twitter account but we have to take into consideration four major parameters for the tweet data gathering – What is the Consumer Key along with Consumer Secret Key also we need the Access Token and Access Token Secret to complete the process.

So to achieve that we collect the tweets through Twitter Streaming API with geo-location filter. The advantage of that is all tweets in this are paired with time and location information. To ensure that the locations are not repeated, we tokenise the location and separate them out using regex string functionality in Python.

The following actions were taken with respect to specific content in the tweets for better understanding of the correlation:

Table 3. Type of content in tweets with specific action.

Type of Content	Action
Emoticon	Remove
Location: New Delhi, India	Change to New Delhi
Location: Jaipur-The Pink City	Change to Jaipur

In our study for the short term analysis of the correlation between twitter and traffic, we took up the case of one of the most popular cricket league in the world - Indian Premier League. Our analysis began with selecting a

particular match of the league, which in this case was between the teams Mumbai Indians and Delhi Daredevils in Feroz Shah Kotla Stadium, New Delhi. The analysis began with generating a hashtag based identification algorithm, that extracted tweets based on hashtags entered, which in this case were the official hashtags of the league - #IPL and #DDvMI. The tweets with no location, or with any of the exceptions mentioned above, are ignored. These tweets are stored in a SQL database with a timestamp and the location for further analysis.

The same step is performed for different time slots within the same day of the event. Along with the tweet information, the traffic information based on estimation of number of vehicles based on the colour code represented by the Traffic Layer of Google Maps, is also stored in the database. The traffic layer represents the amount of traffic on roads in the form of different colours. According to the Google Maps site, the colored stripes represents traffic conditions on major highways refer to the speed at which one can travel on that road. The deep red lines mean highway traffic is moving more or less at the speed less than 25 miles per hour and it clearly is an indication of a traffic jam or an accident related blockage or an unavoidable congestion on that route. Whereas if it shows yellow colored stripes over the map then it signifies traffic is moving faster atleast, from 25 to 50 miles per hour, and if the green stripes are visible in the google map then that indicates the route is good to go and the traffic is moving faster on that route at 50 miles or more per hour. In between we also often see grey lines, that just simply means that on that particular route there's no traffic information available at the time and we may also find the red-black line too visible in mapping app which happens to refers to extremely slow or stopped traffic.

If otherwise we have to analyze the traffic on metro city streets, with limitations of speed on motor vehicles such that the speed has to kept much lower than on the highways, in that case the colors on google map take on different color code relative meaning. In the reddish blacklines denote a slow going traffic along with general congestion most of the time. The yellow lined depict a better traffic prospect but still not the best and fast route for city travel, and the green denotes the traffic conditions are suitable to travel. Based on these facts, the traffic can be estimated, and given certain values for analysis for the event that begins at around evening 8pm. and ends around

late at night 11pm.

By storing values of the number of tweets and the amount of traffic throughout the day of the event, we noticed a trend being followed. The map indicates the traffic condition around the stadium 4 hours before the match begins, i.e., 4:00 P.M. to be quite low. The condition changes drastically after this period, around 7:00 P.M., wherein the number of vehicles increases manifold. Similarly, the number of tweets with #IPL and #DDvMI from New Delhi go up considerably than other places from 6:00 P.M. until 8:00 P.M.; go down till 11:00 P.M., in accordance with the amount of traffic in the region, and as soon as the event ends, both the amount of traffic and tweets shoot up again.

The graph below shows the correlation between traffic with respect to time, with the traffic being estimated as per the colour codes indicated by the Traffic Layer of Google Maps.

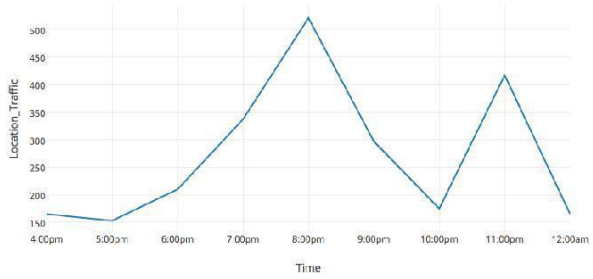


Figure 6. Traffic at Feroz Shah Kotla Stadium, New Delhi, at different time intervals on the day of the event.

V. Conclusion

In the above written research paper, we had been motivated and intrigued by the fact that nowadays more n more persons are in habit to post any public mass gathering event-related contents on popular social media platforms and also always are on the run to and from the place of the event. We answered the following question via this paper: can we utilize such information or tweets to improve traffic prediction for similar events in the future. To prove this aspect, we first performed correlation analysis between posted tweets counts and traffic measurements for a long term basis, by considering the case of Wimbledon - one of the biggest sporting events, and

then work on a short term analysis by considering the example of the Indian Premier League. We analysed the tweets on a hashtag based event identification algorithm, and further used the Google Maps and its layers to predict traffic over a certain period, and then use it for future prediction. Based on the derived analysis, we have come to the conclusion that twitter semantics can surely be used for general traffic prediction for planned events, and hence improve traffic management. Experimental results on traffic data and Twitter data collected for United Kingdom clearly depicted the improved performance of our proposed model based on auto-regression over the existing traffic prediction model.

References

[1] S. Petrović, M. Osborne and V. Lavrenko, Streaming first story detection with application to twitter, In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (2010), (pp. 181-189). Association for Computational Linguistics.

[2] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman and J. Sperling, Twitterstand: news in tweets, In Proceedings of the 17th acm sigspatial international conference on advances in geographic information systems (2009), 42-51. ACM.

[3] T. Sakaki, M. Okazaki and Y. Matsuo, Earthquake shakes Twitter users: real-time event detection by social sensors, In Proceedings of the 19th International Conference on World Wide Web (2010), 851-860. ACM.

[4] A. Go, R. Bhayani and L. Huang, Twitter sentiment classification using distantsupervision, CS224N Project Report, Stanford 1 (2009), 12.

[5] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu and M. Demirbas, Short text classification in twitter to improve information filtering, In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (2010), 841-842. ACM.

[6] J. Bollen, H. Mao and X. Zeng, Twitter mood predicts the stock market, Journal of Computational Science 2(1) (2011), 1-8.

[7] J. Eisenstein, B. O'Connor, N. A. Smith and E. P. Xing, A latent variable model for geographic lexical variation, In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (2010), 1277-1287. Association for Computational Linguistics.

[8] T. Maze, M. Agarwai and G. Burchett, Whether weather matters to traffic demand, traffic safety, and traffic operations and flow, Transportation research record: Journal of the transportation research board (1948), 170-176.

[9] H. S. Mahmassani and J. Dong, Journal of the Transportation Research Board, No. 2391,

Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 56-68. DOI: 10.3141/2391-06

[10] J. Kim, R. B. Chen and B. Park, Incorporating weather impacts in traffic estimation and predication systems, US Department of Transport, Washington (2009), 108.

[11] M. Ni, Q. He and J. Gao, Forecasting the subway passenger flow under event occurrences with social media, IEEE Transactions on Intelligent Transportation Systems 18(6) (2017), 1623-1632.

[12] O. Ozdikis, P. Senkul and H. Oguztuzun, Semantic expansion of hashtags for enhanced event detection in Twitter, In Proceedings of the 1st international workshop on online social systems (2012).

[13] J. He, W. Shen, P. Divakaruni, L. Wynter and R. Lawrence, Improving traffic prediction with tweet semantics, In Twenty-Third International Joint Conference on Artificial Intelligence (2013, June).